

Il valore Shapley come indice di rilevanza in un network

Giulia Cesari

Politecnico di Milano e Université Paris Dauphine
giulia.cesari@polimi.it

Almo Collegio Borromeo, 5 maggio 2017

Misurare la rilevanza di nodi all'interno di un network

- ▶ Il concetto di *centralità* gioca un ruolo chiave in molte applicazioni al mondo reale ed è stato studiato ampiamente in "network analysis".
- ▶ È spesso naturale chiedersi quali siano i nodi o lati più importanti all'interno di un certo network.
- ▶ Questa questione può sorgere in molti contesti differenti: può rivelarsi utile, ad esempio, determinare quali siano:
 - ▶ le persone più influenti in un social network;
 - ▶ le vie principali in una rete stradale;
 - ▶ i geni più importanti in un network di regolazione genica;
 - ▶ gli argomenti più conflittuali all'interno di un discorso.

Previously on "Teoria dei giochi":

- ▶ giochi cooperativi con restrizione sulle possibilità di interazione tra i giocatori: *communication situations* e *graph games*
- ▶ *microarray games* e valore Shapley come indice di rilevanza di geni in una condizione biologica come l'insorgere di una malattia genetica

Tornando ai geni..

- ▶ Un gioco cooperativo descrive una situazione in cui tutti i giocatori possono liberamente interagire tra loro
 - ▶ tutte le coalizioni sono ammissibili
- ▶ Facciamo cadere questa ipotesi:
 - ▶ introduciamo una restrizione sulle possibilità di interazione dei giocatori

- ▶ *Qual è il significato di imporre una restrizione sulle possibilità di interazione tra geni?*
- ▶ *Quali informazioni ci fornisce un network di geni al fine di individuare geni rilevanti all'interno di un contesto biologico?*

..network biologici

- ▶ I meccanismi di interazione tra geni, RNA e proteine sono molto complessi e oggetto di grande interesse nel campo della ricerca biomedica e epidemiologica.
- ▶ Tali meccanismi sono descritti da reti di regolazione genica: *gene regulatory network* o *gene regulatory pathway*.
- ▶ La ricostruzione di tali meccanismi è fondamentale per la comprensione delle funzioni di geni nella determinazione di una certa condizione biologica di interesse, come l'insorgere di una malattia genetica.
- ▶ L'interpretazione dell'interazione tra geni all'interno di network biologici rende dunque necessaria l'individuazione di **misure dell'importanza di geni all'interno di tali network**.

Centralità di geni in un network biologico

- ▶ Diversi approcci sono stati proposti per l'identificazione di geni 'centrali' all'interno di pathway biologici.
- ▶ I **network di coespressione** di geni (gene co-expression network), sono sempre più usati per studiare il ruolo di geni e proteine all'interno dei meccanismi di regolazione che avvengono a livello cellulare.

Network di coespressione

- ▶ i nodi del network rappresentano i geni
- ▶ la connessione tra nodi è determinata dalla coespressione dei geni nei campioni in esame (coefficiente di correlazione di Pearson)

Lo studio della *centralità* di elementi in un network rappresenta un importante strumento per l'interpretazione dell'interazione tra geni all'interno di network di coespressione.

- ▶ Jeong et al. (2001): "Lethality and centrality in protein networks."
- ▶ Carlson et al. (2006): "Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks."

Cosa si intende con "centralità"?

Quali sono possibili misure di centralità di geni in un network biologico?

- ▶ Gli studi di centralità nascono nell'ambito dell'analisi dei [social network](#)

Social network

- ▶ Un social network è una struttura sociale fatta di nodi (generalmente rappresentano individui o organizzazioni) che sono messi in relazione a coppie a rappresentare una o più tipologie di interdipendenza, come:
 - ▶ valori, ideali, scambi finanziari, amicizia, antipatia, conflitto, parentela, commercio...
- ▶ La struttura che ne risulta può essere rappresentata mediante un grafo



Analisi dei social network

- ▶ Il corpus teorico e i modelli usati per lo studio delle reti sociali sono compresi nella cosiddetta *social network analysis*
- ▶ La ricerca condotta nell'ambito di diversi approcci disciplinari ha evidenziato come le reti sociali operino a più livelli (dalle famiglie alle comunità nazionali) e svolgano un ruolo cruciale nel determinare le modalità di risoluzione di problemi e i sistemi di gestione delle organizzazioni, nonché le possibilità dei singoli individui di raggiungere i propri obiettivi
- ▶ La metafora dei social network è stata utilizzata per più di un secolo per rappresentare insieme di complesse interrelazioni tra i membri di un sistema sociale su varie scale, da quelle interpersonali a quelle internazionali

- ▶ Lo studio dei network sociali si è trasformato dall'essere una suggestiva metafora per diventare un approccio analitico vero proprio, con i suoi enunciati teorici, metodi di ricerca specifici e ricercatori specializzati.
- ▶ Vengono utilizzati in diversi settori della scienza applicata: antropologia, biologia, studi di comunicazione, economia, geografia, informatica, organizzazione, psicologia sociale e socio-linguistica

Metodi di misura nei social network

Hanno lo scopo di studiare diverse proprietà:

- ▶ Centralità, coesione
- ▶ Coefficienti di clustering ('cliquishness')
- ▶ Coefficienti strutturali
- ▶ ...

CENTRALITÀ: indicazione del potere sociale dei nodi basato sulla capacità di rendere "connesso" il network

Primi studi sulla centralità

- ▶ L'idea di centralità applicata alla comunicazione tra individui fu introdotta dallo studioso di psicologia americano Bavelas nel 1948.
- ▶ Studi riguardante la comunicazione in piccoli gruppi di individui: ipotesi di relazione tra centralità strutturale e influenza all'interno di processi di gruppo.
- ▶ Le ricerche evidenziarono legame tra centralità e efficienza del gruppo in processi di problem-solving, percezione della leadership e personale soddisfazione dei partecipanti al gruppo

Freeman - "*Centrality in social networks: conceptual clarification*"(1979)

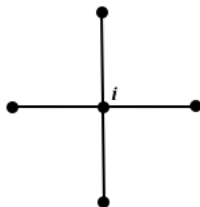
Il sociologo L. Freeman scrive, nel 1979: *“L’idea di centralità è viva ed è stata mobilitata in una varietà di applicazioni sempre più ampia. Tutti concordano, a quanto pare, sul fatto che la centralità sia un importante attributo della struttura di un network sociale. [...] Ma qui il consenso finisce. Non vi è sicuramente accordo su cosa sia esattamente la centralità o sulle sue radici concettuali, e vi è poco consenso sulle corrette procedure per la sua misurazione.”*

Centralità in un network: misure classiche

Idea di base: il fulcro di una stella (*hub*) è la posizione più centrale possibile.

Cos'è che rende unica la posizione di i ?

- ▶ i può comunicare con molti altri nodi;
- ▶ vi sono molti altri nodi che necessitano di i come intermediario nelle loro comunicazioni;
- ▶ i è vicino a molti altri nodi.

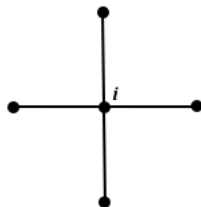


Centralità in un network: misure classiche

Idea di base: il fulcro di una stella (*hub*) è la posizione più centrale possibile.

Cos'è che rende unica la posizione di i ?

- ▶ possiede il massimo grado possibile (*degree centrality*)
- ▶ giace sul massimo numero di geodetiche che collegano gli altri vertici (*betweenness centrality*)
- ▶ si trova alla minima distanza da tutti gli altri vertici (*closeness centrality*)



Degree centrality

Degree centrality (Nieminen, 1974; Shaw, 1954): la degree centrality di un nodo è definita come il numero di lati incidenti in esso.

La degree centrality è un indice della potenziale *attività di comunicazione* di un nodo: maggiore la possibilità di comunicare direttamente con gli altri nodi, maggiore la centralità.

Betweenness centrality

Betweenness centrality (Bavelas, 1948; Freeman, 1977): la betweenness centrality di un nodo k è definita come $\sum_{i,j \in N \setminus \{k\}} b_{ij}(k)$, dove $b_{ij}(k) = \frac{g_{ij}(k)}{g_{ij}}$, dove g_{ij} è il numero di geodetiche che collegano i nodi i e j , mentre $g_{ij}(k)$ è il numero di geodetiche che collegano i nodi i e j e contengono k .

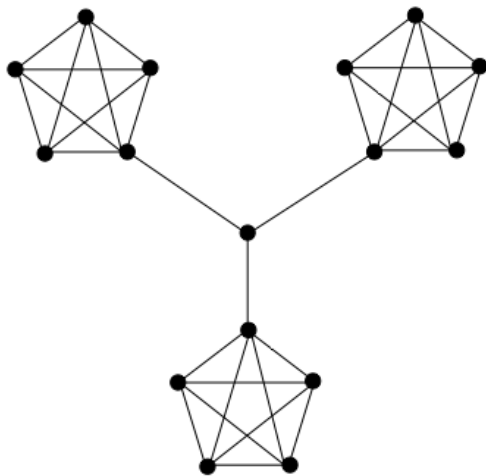
La betweenness centrality è un indice del potenziale di un nodo nel *controllo della comunicazione*: maggiore il ruolo di intermediario nelle comunicazioni tra gli altri nodi, maggiore la centralità.

Closeness centrality

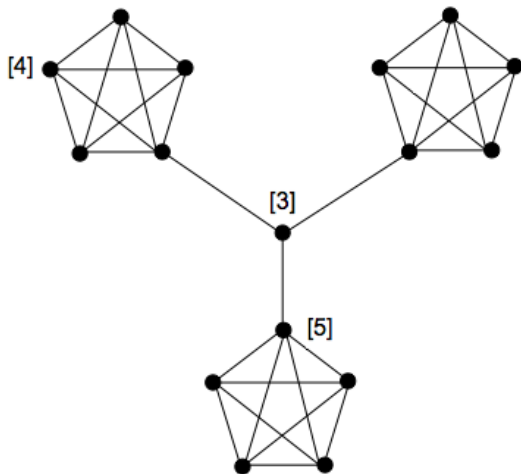
Closeness centrality (Beauchamp, 1965; Sabidussi, 1966): la closeness centrality di un nodo i è definita come $\frac{|N|-1}{\sum_{j \in N} d(i,j)}$, dove $d(i,j)$ è la distanza tra i nodi i e j , ossia la lunghezza del cammino più breve tra i e j .

La closeness centrality è anch'essa collegata al controllo della comunicazione, ma in maniera diversa. Un nodo risulta centrale in questo caso nella misura in cui può evitare il potenziale di controllo degli altri nodi. La closeness centrality è un indice dell' **indipendenza di un nodo**: la possibilità di comunicare con molti altri nodi dipendendo da un numero minimo di intermediari.

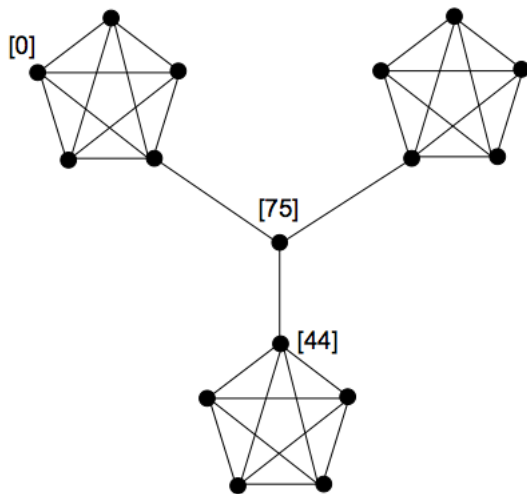
Esempio: qual è il nodo più “centrale”?



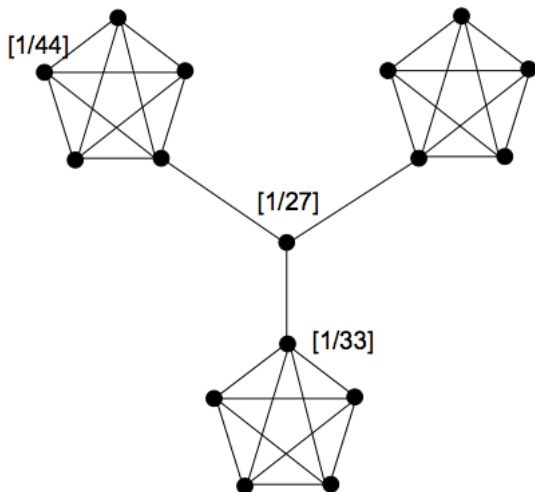
Esempio: degree centrality



Esempio: betweenness centrality



Esempio: closeness centrality



Misure di centralità classiche

Le misure di centralità classiche assegnano ad ogni nodo di un network un valore che corrisponde in qualche modo all'importanza di tale nodo per l'applicazione in esame.

Esempio

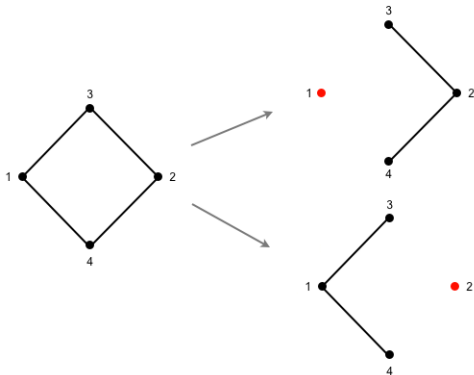
Nella progettazione di un network di infrastrutture che sia il meno possibile vulnerabile al guasto di un nodo, **una misura di centralità classica potrebbe assegnare un valore ad ogni nodo in modo proporzionale ai danni conseguenti dal suo guasto.**

Misure di centralità classiche: limitazioni

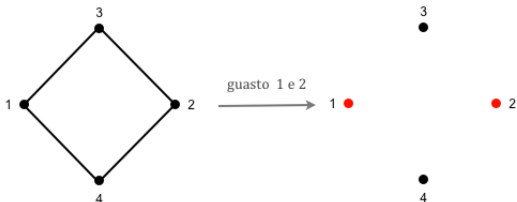
- ▶ Poichè i nodi vengono valutati separatamente, vi è l'assunzione implicita che i guasti dei nodi avvengano indipendentemente gli uni dagli altri. Di conseguenza, fenomeni comuni come i guasti di nodi a cascata sfuggono ad una tale analisi.
- ▶ Considerando solamente il guasto di nodi singoli, si ignorano situazioni più realistiche in cui diversi nodi possono venire meno contemporaneamente.

Le misure di centralità classiche possono rivelarsi inefficaci nel riflettere il ruolo delle coppie di nodi o più in generale di sottoinsiemi di nodi.

Esempio: rete di comunicazione (guasto di un nodo)



Esempio: rete di comunicazione (guasto di una coppia di nodi)



- ▶ Misure di centralità classiche non sono in grado di riconoscere che in molte applicazioni reali non è sufficiente considerare i nodi come entità a sé stanti.
- ▶ Un requisito importante consiste nella comprensione dell'importanza di ciascun nodo in termini della sua utilità congiunta con altri nodi.

- ▶ Misure di centralità classiche non sono in grado di riconoscere che in molte applicazioni reali non è sufficiente considerare i nodi come entità a sé stanti.
- ▶ Un requisito importante consiste nella comprensione dell'importanza di ciascun nodo in termini della sua utilità congiunta con altri nodi.

Game theoretic network centrality

Un approccio all'analisi di centralità con la Teoria dei giochi (*game theoretic network centrality*) è stato proposto come quadro teorico per affrontare tali limitazioni.

Teoria dei giochi e centralità in un network

- ▶ Idea di base: definire un gioco cooperativo in cui i giocatori siano i nodi del network in esame.
- ▶ L'indice di potere di un nodo è interpretato come misura di centralità, in quanto rappresenta il contributo marginale atteso di ogni nodo ad ogni possibile combinazione di altri nodi

Game theoretic network centrality

- ▶ Gómez et al. (2003): "Centrality and power in social networks: A game theoretic approach."
- ▶ Suri e Narahari (2008): "Determining the top-k nodes in social networks using the Shapley Value."

Teoria dei giochi in un network sociale

Gómez et al. (2003)

- ▶ La centralità di un individuo viene misurata come variazione di potere dovuta alla struttura sociale a cui appartiene (i.e. alla restrizione delle possibilità di comunicazione imposta dal network).
- ▶ Si basa sui giochi ristretti al grafo e sul valore Myerson.

Misura di centralità γ

$$\gamma_i(v, \Gamma) = \phi_i(v^\Gamma) - \phi_i(v) \quad \text{per ogni } i \in N$$

Proprietà di γ

Se ν è simmetrico e convesso:

- ▶ Equità : la rimozione di un lato produce la stessa variazione di centralità per entrambi i nodi incidenti
- ▶ Stabilità: la rimozione di un lato causa una diminuzione di centralità per entrambi i nodi incidenti
- ▶ La centralità di un nodo all'interno del grafo è pari alla centralità nel sottografo indotto dalla componente connessa a cui appartiene (cfr. *degree*)
- ▶ I nodi isolati hanno centralità minima (cfr. *degree*)
- ▶ Tra tutti i grafi con n nodi, la massima centralità è assegnata al fulcro di una stella (cfr. *degree, betweenness, closeness*)
- ▶ Tra tutti i grafi connessi con n , la minore centralità è assegnata ai nodi estremali di una catena (cfr. *degree, betweenness*)
- ▶ In una catena, la centralità cresce dai nodi estremali verso i nodi mediani (cfr. *degree, betweenness, closeness*)

Top-k nodes problem

Suri and Narahari (2008)

- ▶ Contesto: co-authorship network, diffusione dell'informazione, marketing virale etc.
- ▶ Top k-nodes problem è NP-hard.
- ▶ Il gioco: il valore di una coalizione è il numero di vicini
- ▶ I top-k nodes: i nodi con il valore Shapley più alto

Aadithya et al. (2010)

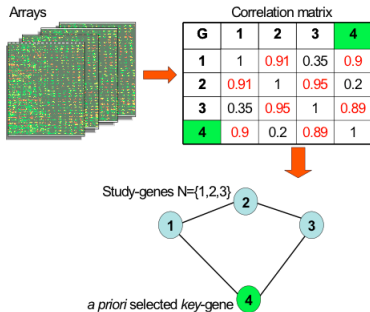
Formula esatta per il valore Shapley del gioco introdotto da Suri and Narahari (2008):

$$\phi_i(N, v, \Gamma) = \sum_{j \in N_i(E)} \frac{1}{1 + \text{deg}(j)}$$

Biologia e network

Diversi modelli basati su network sono utilizzati in biologia molecolare:

- ▶ protein interaction networks
- ▶ gene regulatory networks
- ▶ gene co-expression networks



Interazioni tra geni e centralità

- ▶ Misure di centralità classiche sono appropriate nell'ipotesi che i nodi siano indipendenti gli uni dagli altri e che il sistema sia sensibile alla variazione di espressione di ogni singolo nodo.
- ▶ Al contrario, all'interno dei complessi network biologici, assumere che l'espressione di un gene/proteina sia indipendente da quella degli altri geni/proteine non è realistico e le conseguenze sul sistema si possono apprezzare solo se molti nodi variano la propria espressione.

TdG e centralità in un network biologico (1)

Problema: individuazione di geni rilevanti all'interno di una condizione biologica, tipicamente l'insorgenza di una malattia.

Moretti et al. (2010)

Network di coespressione come situazione di comunicazione:

- ▶ Network $\Gamma = (N, E)$, dove N sono i geni e E le relazioni di coespressione
- ▶ (N, v) è un gioco a priori, detto *association game*, dove per ogni $S \subseteq N$, $v(S)$ è il numero di *geni chiave* che interagiscono solo con geni in S .
- ▶ (N, v^Γ) è il **gioco ristretto al grafo** associato.

TdG e centralità in un network biologico (1)

Misura di centralità γ

La rilevanza di un gene $i \in N$ è misurata da:

$$\gamma_i(N, v, \Gamma) = \phi_i(v^\Gamma) - \phi_i(v).$$

- ▶ L'indice di rilevanza può essere calcolato attraverso algoritmi di approssimazione.
- ▶ Un gene è rilevante se è altamente associato alla condizione biologica in esame.

TdG e centralità in un network biologico (2)

Problema: individuazione di geni rilevanti all'interno di una condizione biologica, tipicamente l'insorgenza di una malattia (eventualmente senza la conoscenza a priori di geni chiave all'interno di tale network)

Cesari et al. (2016)

Graph game sul network di co-espressione:

- ▶ $k \in \mathbb{R}^N$ specifica l'importanza a priori di ogni gene
- ▶ (N, v_E^k) gioco definito a partire dal grafo $v_E^k(S) = \sum_{j \in S \cup N_S(E)} k_j$ (estensione del gioco introdotto da Suri e Narahari (2008))
- ▶ **N.B.** È facile osservare che v_E^k è un **basic GAG**

TdG e centralità in un network biologico (2)

Valore Shapley come indice di rilevanza

La rilevanza di un gene $i \in N$ è misurata da:

$$\rho_i(v_E^k) = \sum_{j \in (N_i(E) \cup \{i\})} \frac{k_j}{d_j(E) + 1}.$$

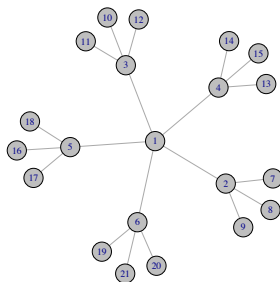
- ▶ L'indice di rilevanza può essere calcolato in maniera esatta grazie ai risultati teorici sulla classe dei basic GAG;
- ▶ ed ha una caratterizzazione assiomatica in cui le proprietà hanno un'interpretazione biologica.

TdG e centralità in un network biologico (2)

- ▶ Un gene è rilevante se ha un ruolo importante nel preservare l'attività di regolazione all'interno del network e come intermediario tra gli hub e i geni più periferici.

I geni più rilevanti sono i nodi 2, 3, 4, 5, 6:

- ▶ grazie alle loro connessioni, sono in grado di influenzare l'espressione di tanti altri nodi;
- ▶ se rimossi (o inibiti), compromettono l'attività di regolazione, lasciando molti nodi isolati o in piccole componenti connesse.

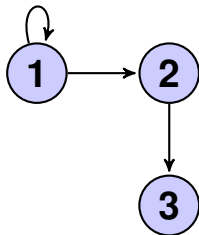


TdG e centralità in un grafo di argomentazione

- ▶ Teoria dell'argomentazione: formalizzazione di sistemi e processi decisionali
- ▶ Obiettivo: ricerca di insiemi di argomenti accettabili

Un **grafo di argomentazione** è un grafo diretto:

- ▶ nodi sono argomenti
- ▶ lati rappresentano attacchi tra coppie di argomenti



Semantiche e accettabilità di argomenti

Semantiche sono criteri per determinare quali argomenti sono accettabili in un contesto di argomentazione (Dung et al.1995, Caminada et al. 2008).

- ▶ La Teoria dei Giochi è stata utilizzata di recente per definire semantiche *graduali*
- ▶ L'*accettabilità* non è l'unico attributo studiato da un punto di vista graduale

Obiettivo: ordinare gli argomenti sulla base della loro *controversialità* attraverso approccio con la Teoria dei Giochi.

Il valore Shapley come indice di conflitto

Amgoud 2015: un approccio assiomatico alla misura del *disagreement* in un grafo di argomentazione.

- ▶ Definiamo un gioco cooperativo sul grafo di argomentazione, in cui il valore di una coalizione è una misura del disaccordo totale nel grafo indotto:

$$v(S) = \frac{\max - D(S)}{\max - \min},$$

dove $D(S) = \sum_{i,j \in S} d_{i,j}$, $\max = n^2(n+1)$ e $\min = n^2$.

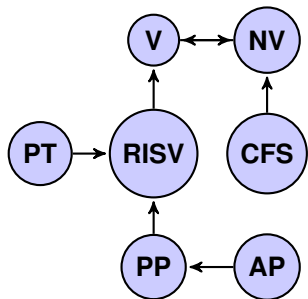
- ▶ Il valore Shapley di tale gioco misura il contributo dei singoli argomenti al disaccordo totale all'interno del grafo ed è interpretato come *indice di conflitto* che misura la controversialità di ogni argomento.

Esempio: un dibattito sul trapianto di un organo

Un grafo di argomentazione con 7 argomenti che rappresenta un dibattito tra due coordinatori responsabili di un trapianto:

- ▶ NV: l'organo non è adatto al trapianto;
- ▶ V: l'organo è adatto al trapianto;
- ▶ RISV: il donatore è affetto da endocardite dovuta allo streptococcus viridans;
- ▶ CFS: l'organo ha struttura e funzionalità e strutture corrette;
- ▶ PP: un'eventuale infezione del paziente può essere prevenuta con un trattamento a base di penicillina;
- ▶ PT: un'eventuale infezione del paziente può essere prevenuta con un trattamento a base di teicoplanina;
- ▶ AP: il paziente è allergico alla penicillina.

Esempio: quali sono gli argomenti più controversi?

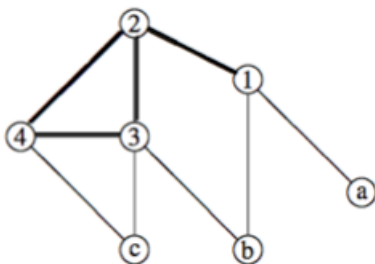


Ranking	argomento	controversialità
1	V	0.0816
2	NV	0.0772
3	RISV	0.0481
4	PP	0.0364
5	AP	0.0321
6	PT	0.0262
7	CFS	0.0190





Esercizio

Per il grafo in figura, dove $N = \{1, 2, 3, 4\}$,
 $E = \{\{1, 2\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$ e a, b e c sono i geni chiave,
calcolare:

- ▶ degree centrality, closeness centrality, betweenness centrality;
- ▶ la misura di centralità γ ;
- ▶ l'indice di rilevanza ρ del gioco v_E^k , dove $k_i = 1$ per ogni $i \in N$.







Bibliografia

-  L. Amgoud, Measuring disagreements in argumentation graphs, Mimeo (2015).
-  L. Amgoud, J. Ben-Naim, Ranking-based semantics for argumentation frameworks, Scalable Uncertainty Management 134-147, Springer Berlin Heidelberg (2013).
-  A. Bavelas: “A mathematical model for group structures.”, *Human Organization*, **7**, 16-30 (1948).
-  G. Cesari, E. Algaba, S. Moretti, J.A. Nepomuceno: “A game theoretic neighborhood-based relevance index to evaluate nodes in gene co-expression networks”, Submitted to PLOS in 2016.





Bibliografia

-  G. Cesari, F. Fossati, S. Moretti, “A conflict index for arguments in an argumentation graph”, Submitted to the 2nd European Conference on Argumentation (ECA2017).
-  L.C. Freeman: “Centrality in social networks: conceptual clarification.”, *Social Networks*, **1**, 215-239, (1979).
-  D. Koschützki, K.A. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl, O. Zlotowski: “Centrality indices. Network analysis.”, *LNCS*, **3418**, 16-61 (2005).
-  D. Gómez, E. González-Arangüena, C. Manuel, G. Owen, M. Del Pozo, J. Tejada: “Centrality and power in social networks: a game theoretic approach.”, *Mathematical Social Sciences*, **46**, 27-54 (2003).

Bibliografia

-  N. R. Suri, Y. Narahari: “Determining the top-k nodes in social networks using the Shapley Value.”, *AAMAS 2008*, 1509-1512 (2008).
-  N. R. Suri, Y. Narahari: “A Shapley Value Based Approach to Discover Influential Nodes in Social Networks.”, *IEEE 2011*, **8**, 130-147 (2011).
-  K. V. Aadithya, B. Ravindran: “Game theoretic network centrality: exact formulas and efficient algorithms”, *AAMAS*, (2010).
-  K. V. Aadithya, B. Ravindran, T.P. Michalak, N.R. Jennings: “Efficient computation of the Shapley Value for centrality in networks.”, *WINE 2010, LNCS*, **6484**, 1-13 (2010).

Bibliografia

-  B. Zhang, S. Horvath: “A general framework for weighted gene co-expression network analysis.”, *Statistical Applications in Genetics and Molecular Biology*, **4**, (2005).
-  H. Jeong, S.P. Mason, A.-L. Barabási: “Lethality and centrality in protein networks.”, *Nature*, **411**, 41-42 (2001).
-  M. RJ Carlson et al.: “Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks.”, *BMC Genomics*, **7**, 40 (2006).
-  S. Moretti, V. Fragnelli, F. Patrone, S. Bonassi: “Using coalitional games on biological networks to measure centrality and power of genes.”, *Bioinformatics*, **26**, 2721-2730 (2010).